CLAIMS

What is claimed is:

1. A method comprising:

processing a plurality of information obtained from a base calling system;

creating a plurality of refined base calls using a plurality of original base calls and a plurality of intrinsic peak characteristics; and

assigning a quality value to each of the plurality of refined base calls using the plurality of intrinsic peak characteristics.

2. The method of claim 1 wherein processing comprises:

detecting a plurality of peaks;

expanding the plurality of peaks into a plurality of expanded peaks; and

resolving the plurality of expanded peaks.

3. The method of claim 2 wherein detecting comprises:

identifying a plurality of inflection points; and

selecting the plurality of peaks based on the plurality of inflection points.

4. The method of Claim 3 wherein detecting further comprises:

computing a plurality of apparent peak characteristics.

5. The method of claim 3 wherein selecting comprises:

ignoring those peaks that are insufficiently large.

6. The method of claim 2 wherein expanding comprises:

scanning to the left from a left inflection point of each peak of the plurality of peaks to determine an expanded left boundary for the peak; and

scanning to the right from a right inflection point of each peak of the plurality of peaks to determine an expanded right boundary for the peak.

7. The method of claim 6 wherein scanning to the left comprises:

if a zero signal is detected, designating the location at which the zero signal is detected as left expanded boundary for the peak;

if the beginning of a trace is detected, designating the location at which the beginning of a trace is detected as the left expanded boundary for the peak;

if a local minimum is detected, designating the location at which the local minimum is detected as the left expanded boundary for the peak;

if a right inflection point of a previous peak is detected, designating the location at which the right inflection point of the previous peak is detected as the left expanded boundary for the peak;

if the left expanded boundary is to the left of a right expanded boundary of the previous peak, redefining the left expanded boundary of the peak and right expanded boundary of the previous peak as the midpoint between a left inflection point of the peak and the right inflection point of the previous peak.

8. The method of claim 6 wherein scanning to the right comprises:

if a zero signal is detected, designating the location at which the zero signal is detected as the right expanded boundary for the peak;

if the end of a trace is detected, designating the location at which the end of the trace is detected as the right expanded boundary for the peak;

if a local minimum is detected, designating the location at which the local minimum is detected as the right expanded boundary for the peak;

if a left inflection point of a next peak is detected, designating the location at which the left inflection point of the next peak is detected as the right expanded boundary for the peak;

if the right expanded boundary is to the right of a left expanded boundary of the next peak, redefining the right expanded boundary of the peak and left expanded boundary of the next peak as the midpoint between a right inflection point of the peak and the left inflection point of the next peak.

9. The method of claim 6 wherein expanding further comprises:

classifying each of the expanded peaks based on the type of expanded left boundary and the type of expanded right boundary.

10. The method of claim 2 wherein resolving comprises:

fitting the plurality of expanded peaks using a model of a peak shape; and

computing a peak resolution.

11. The method of claim 10 wherein fitting comprises:

deriving an equation representing an output of an electrophoresis of a plurality of DNA fragments, the equation accounting for electromigration and diffusion;

solving the equation using a rectangular peak of finite width as an initial condition to obtain a peak shape expression dependent on three adjustable parameters;

for each single peak of the plurality of peaks, computing the plurality of intrinsic peak characteristics using the peak shape expression; and

for each multiple peak portion of the plurality of peaks,

breaking the multiple peak portion into at least two descendent peaks; and

computing the plurality of intrinsic peak characteristics for each of the descendent peaks using the peak shape expression.

12. The method of claim 11 wherein breaking comprises:

computing a set of average peak characteristics based on analysis of a set of preceding single peaks;

calculating a first peak shape parameter and a second peak shape parameter based on the set of average peak characteristics;

for each of the descendent peaks, iteratively computing an intrinsic peak position and an initial height based on the first peak shape parameter, the second peak shape parameter, and the average peak characteristics.

13. The method of claim 10 wherein computing a peak resolution comprises:

dividing an area under a curve representing the absolute value of the difference between an apparent peak signal of an apparent peak of the plurality of peaks and a corresponding fitted model peak by an area of the apparent peak.

14. The method of claim 1 wherein creating comprises:

refining the plurality of original base calls using the plurality of intrinsic peak characteristics;

inserting a plurality of newly called bases using the plurality of intrinsic peak characteristics.

15. The method of claim 14 wherein refining comprises:

calling true peaks;

resolving wide peaks;

re-calling unknown bases; and

removing unmatched bases.

16. The method of claim 14 wherein refining comprises:

for each known original base call,

scanning a peak list of a particular color base;

if a peak is found at a location of the original base call and the peak is not assigned, calling the peak and assigning a corresponding base to the peak;

if the peak is found at the location of the original base call and the peak has been assigned, determining whether the peak is wide enough to be split;

if the peak is wide enough to be split, resolving the peak into two peaks;

if the peak is not wide enough to be split and no good peaks are located near the found peak, calling the peak and assigning a corresponding base to the peak; and

if the peak is not wide enough to be split and a good peak is located near the found peak, converting the found base call to unknown.

17. The method of claim 14 wherein refining comprises:

for each unknown base call,

scanning four peak lists, one peak list for each base;

if a peak is found at a location of an unknown base call, obtaining a best peak;

if the best peak is not assigned, calling the best peak and assigning a corresponding base to the best peak;

if the best peak has been assigned, determining whether the best peak is wide enough to be split;

if the best peak is wide enough to be split, resolving the best peak into two peaks, calling the two peaks, and assigning corresponding bases to the two peaks;

if the best peak is not wide enough to be split, assigning the corresponding base to the best peak;

if no peaks are found at the location of the unknown base call, locating a good peak near the location of the unknown base call;

if no good peak is located near the location of the unknown base call, rejecting the unknown base call; and

if the good peak is located near the location of the unknown base call, calling the good peak and assigning the corresponding base to the peak.

18. The method of claim 11 wherein inserting the plurality of newly called bases comprises:

creating a multi-colored list of all peaks; and

calling those peaks in the multi-colored list of all peaks that meet all of a plurality of criteria.

19. The method of claim 18 wherein the plurality of criteria comprise:

whether an index of an uncalled peak is greater than a specified minimum;

whether an intrinsic height of the uncalled peak is greater than an intrinsic signal of any other peaks at the position of the uncalled peak; and

whether a spacing between two adjacent called peaks is large enough for insertion of a new base.

20. The method of claim 1 wherein assigning comprises:

computing a plurality of peak trace parameters for each intrinsic peak in the plurality of refined base calls; and

obtaining a quality value for each base in the plurality of refined base calls from a look-up table.

21. The method of claim 20 wherein the plurality of peak trace parameters comprise:

a first peak height ratio for a current peak based on a first plurality of called peaks centered at a current peak;

a second peak height ratio for the current peak based on a second plurality of called peaks centered at the current peak;

a peak spacing ratio for the current peak based on the second plurality of called peaks centered at the current peak; and

a peak resolution.

22. The method of claim 21 wherein the first plurality of called peaks centered at the current peak comprise three called peaks, and wherein the second plurality of called peaks centered at the current peak comprise seven called peaks.

23. The method of claim 20 wherein obtaining comprises:

selecting the quality value from the look-up table from the row in the look-up table in which a plurality of table trace parameters in the row of the look-up table each exceed the plurality of peak trace parameters for a current peak.

24. The method of claim 1 further comprising:

training the system by generating a training file and a look-up table.

25. The method of claim 24 wherein training is conducted at each of a plurality of user locations based on a plurality of characteristics and a plurality of requirements of each of the plurality of user locations.

26. A machine readable medium having stored thereon instructions which when executed by a processor cause the machine to perform operations comprising:

processing a plurality of information obtained from a base calling system;

creating a plurality of refined base calls using a plurality of original base calls and a plurality of intrinsic peak characteristics; and

assigning a quality value to each of the plurality of refined base calls using the plurality of intrinsic peak characteristics.

27. The machine readable medium of claim 26 wherein processing comprises:

detecting a plurality of peaks;

expanding the plurality of peaks into a plurality of expanded peaks; and

resolving the plurality of expanded peaks.

28. The machine readable medium of claim 27 wherein detecting comprises:
identifying a plurality of inflection points; and
selecting the plurality of peaks based on the plurality of inflection points.

29. The machine readable medium of Claim 28 wherein detecting further comprises:
computing a plurality of apparent peak characteristics.

30. The machine readable medium of claim 27 wherein expanding comprises:
scanning to the left from a left inflection point of each peak of the plurality of peaks to determine an expanded left boundary for the peak; and
scanning to the right from a right inflection point of each peak of the plurality of peaks to determine an expanded right boundary for the peak.

31. The machine readable medium of claim 27 wherein resolving comprises:
fitting the plurality of expanded peaks using a model of a peak shape; and
computing a peak resolution.

32. The machine readable medium of claim 31 wherein fitting comprises:
deriving an equation representing an output of an electrophoresis of a plurality of DNA fragments, the equation accounting for electromigration and diffusion;
solving the equation using a rectangular peak of finite width as an initial condition to obtain a peak shape expression dependent on a set of adjustable parameters;
for each single peak of the plurality of peaks, computing the plurality of intrinsic peak characteristics using the peak shape expression; and
for each multiple peak portion of the plurality of peaks,
breaking the multiple peak portion into at least two descendent peaks; and
computing the plurality of intrinsic peak characteristics for each of the descendent peaks using the peak shape expression.

33. The machine readable medium of claim 32 wherein breaking comprises:
computing a set of average peak characteristics based on analysis of a set of preceding single peaks;
calculating a first peak shape parameter and a second peak shape parameter based on the set of average peak characteristics;

for each of the descendent peaks, iteratively computing an intrinsic peak position and an initial height based on the first peak shape parameter, the second peak shape parameter, and the average peak characteristics.

34. The machine readable medium of claim 31 wherein computing the peak resolution comprises:

dividing an area under a curve representing the absolute value of the difference between an apparent peak signal of an apparent peak of the plurality of peaks and a corresponding fitted model peak by an area of the apparent peak.

35. The machine readable medium of claim 26 wherein creating comprises:

refining the plurality of original base calls using the plurality of intrinsic peak characteristics;

inserting a plurality of newly called bases using the plurality of intrinsic peak characteristics.

36. The machine readable medium of claim 35 wherein refining comprises:

calling true peaks;

resolving wide peaks;

re-calling unknown bases; and

removing unmatched bases.

37. The machine readable medium of claim 35 wherein refining comprises:

for each known original base call,

scanning a peak list of a particular color base;

if a peak is found at a location of the original base call and the peak is not assigned, calling the peak and assigning a corresponding base to the peak;

if the peak is found at the location of the original base call and the peak has been assigned, determining whether the peak is wide enough to be split;

if the peak is wide enough to be split, resolving the peak into two peaks;

if the peak is not wide enough to be split and no good peaks are located near the found peak, calling the peak and assigning a corresponding base to the peak; and

if the peak is not wide enough to be split and a good peak is located near the found peak, converting the found base call to unknown.

38. The machine readable medium of claim 35 wherein refining comprises:

for each unknown base call,

scanning four peak lists, one peak list for each base;

if a peak is found at a location of an unknown base call, obtaining a best peak;

if the best peak is not assigned, calling the best peak and assigning a corresponding base to the best peak;

if the best peak has been assigned, determining whether the best peak is wide enough to be split;

if the best peak is wide enough to be split, resolving the best peak into two peaks, calling the two peaks, and assigning corresponding bases to the two peaks;

if the best peak is not wide enough to be split, assigning the corresponding base to the best peak;

if no peaks are found at the location of the unknown base call, locating a good peak near the location of the unknown base call;

if no good peak is located near the location of the unknown base call, rejecting the unknown base call; and

if the good peak is located near the location of the unknown base call, calling the good peak and assigning the corresponding base to the peak.

39. The machine readable medium of claim 35 wherein inserting the plurality of newly called bases comprises:

creating a multi-colored list of all peaks; and

calling those peaks in the multi-colored list of all peaks that meet all of a plurality of criteria.

40. The machine readable medium of claim 39 wherein the plurality of criteria comprise:

whether an index of an uncalled peak is greater than a specified minimum;

whether an intrinsic height of the uncalled peak is greater than an intrinsic signal of any other peaks at the position of the uncalled peak; and

whether a spacing between two adjacent called peaks is large enough for insertion of a new base.

41. The machine readable medium of claim 26 wherein assigning comprises:

computing a plurality of peak trace parameters for each intrinsic peak in the plurality of refined base calls; and

obtaining a quality value for each base in the plurality of refined base calls from a look-up table.

42. The machine readable medium of claim 41 wherein the plurality of peak trace parameters comprise:

a first peak height ratio for a current peak based on a first plurality of called peaks centered at a current peak;

a second peak height ratio for the current peak based on a second plurality of called peaks centered at the current peak;

a peak spacing ratio for the current peak based on the second plurality of called peaks centered at the current peak; and

a peak resolution.

43. The machine readable medium of claim 41 wherein obtaining comprises:

selecting the quality value from the look-up table from the row in the look-up table in which a plurality of table trace parameters in the row of the look-up table each exceed the plurality of peak trace parameters for a current peak.

44. The machine readable medium of claim 26 containing further instructions which when executed by a processor cause the machine to perform operations comprising:

training the system by generating a training file and a look-up table.